The following supplements accompany the article

A new method to identify robust climate analogues

Carsten Walther*, Matthias Lüdeke, Ramana Gudipudi

*Corresponding author: carsten.walther@pik-potsdam.de

Climate Research 78: 179–187 (2019)

Supplement

Suppl. 1: Choice of clustering method

The definition of robust climate analogues implies that the appropriate clustering method should emphasize the cluster properties of compactness and distance over connectedness (Janssen et al. 2012). Therefore methods which tend to generate spherical clusters seem adequate. As the analysis of the topological structure and potential dimension reduction (by SOM-based methods, e.g. Kohonen 1998) are not of major importance in our case the partitioning cluster method k-means (MacQueen 1967; Hartigan and Wong 1979) has been applied. This algorithm minimizes the total within-cluster sum-of-squares (*TSS*) criterion (Steinley 2006). If the data set consists of *V* variables and the number of groups is chosen to be *K*, the criterion is defined by:

$$TSS = \sum_{j=1}^{V} \sum_{k=1}^{K} \sum_{o \in Q_k} (x_{oj} - \bar{x}_j^{(k)})^2, \quad (1)$$

where x_{oj} is the value of the variable *j* of object *o* and $\bar{x}_j^{(k)}$ is the mean value of variable *j* in cluster *k*. In our case *o* counts all 3**N* objects in climate space defined by grid element *i* and index $d \in \{c, l, u\}$while Q_k denotes the objects belonging to cluster *k*. A hierarchical clustering is used to initialize the partitioning method. The objects are assigned to the given *k* initial cluster centers followed by a calculation of new centers as the average of all objects within each cluster. In an iterative process each object is assigned to a cluster in the way that *TSS* is minimized again followed by a calculation of the centers. This process is repeated until a breakup criterion is reached.

Suppl. 2: Selection of number of clusters

A challenge in applying k-means is the question of the appropriate number of clusters. Here a consistency measure is applied (Sietz et al. 2011) which identifies the most robust partition. This is done by repeating the clustering algorithm many times with different initial conditions for a sequence of fixed cluster numbers and analyzing the variation in the partitioning results for a given cluster number. The cluster number which – independent of the initial condition – repeatedly generates a similar cluster partition (measured by the Rand-index; Hubert & Arabie 1985) is assumed to be the most appropriate to the given data structure. The resulting consistency measure may show local maxima for different cluster numbers. Amongst these, the appropriate cluster number can be chosen according to a further criteria, e.g. the expected detail of the analysis (Janssen et al. 2012).

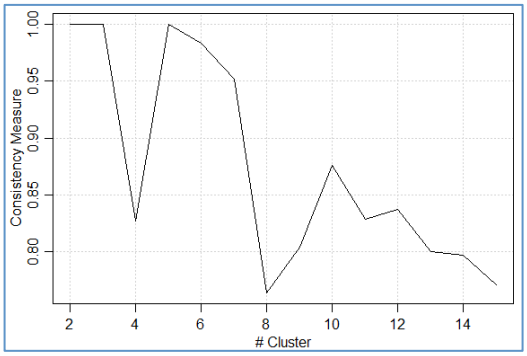

Fig. S1: Result of the consistency measure calculation for the cluster numbers 2 to 15. Pronounced relative maxima denote cluster numbers which reflect the data structure.

For this study we assembled the climate variables in each grid cell from the three climate data sets: the current, the lower and the upper bound of the projections. Then the consistency measure calculation for a varying number of clusters (starting from two clusters and ending at fifteen, see Figure S1) was done. The measure shows pronounced relative maxima at 2, 3, 5 and 10 clusters which reflect the underlying structure of the point cloud in climate space. Climate analogues can only fulfill their purpose with an adequate level of differentiation – therefore cluster numbers 2, 3 and 5 are too small and the partition with ten climate analogue clusters will be used for further analysis.
After performing the selection of the most appropriate cluster number *K* and the clustering into the *K* clusters in climate space the results can be transformed into spatial maps of cluster membership. To further interpret and discuss these clusters, graphs of the distribution of the climate variables in each cluster were developed.

Suppl. 3: Feature plot of the ten climate analogue clusters

For further characterization of the clusters we show in the following Figure S2 the cluster-specific distribution of some selected climate variables.

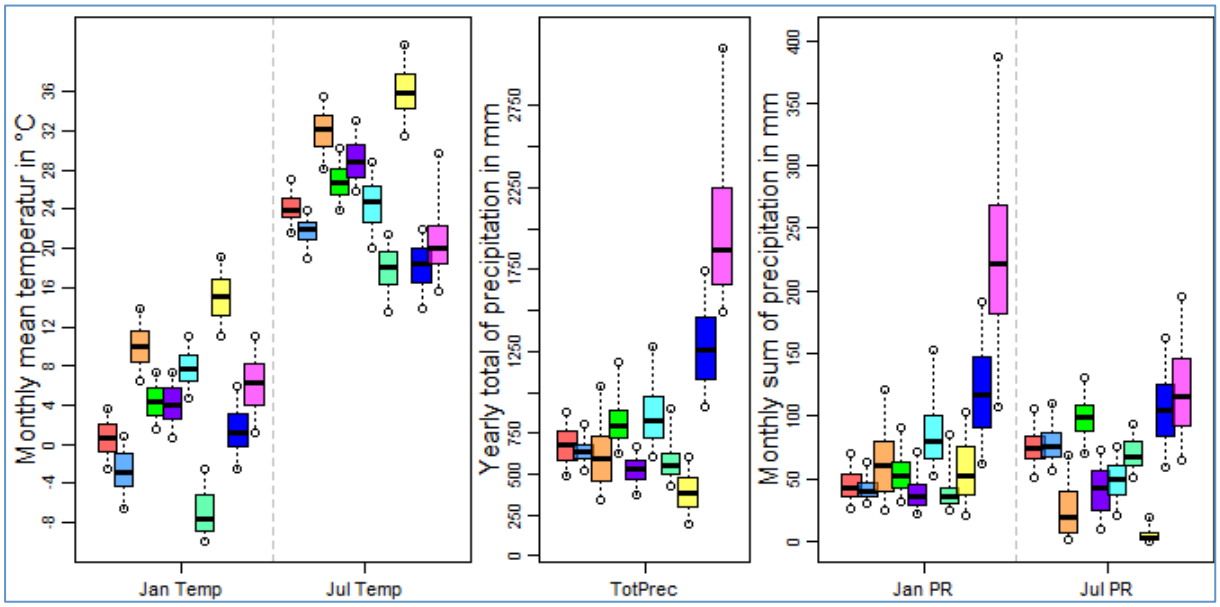

Fig. S2: Feature plot of the ten climate analogue clusters (see e.g. Kok et al. 2016). Examples are shown for monthly mean temperature in the left plot, for total precipitation in the middle (Yearly total – TotPrec) and monthly sum of precipitation in the right plot (January – Jan PR, July – Jul PR). The distribution of the grid cells in the clusters is illustrated by box plots (median, box: 25th and 75th percentile, whiskers: 5th and 95th percentile; the cluster number corresponds to the order of the boxes in the plot from left to right).

Suppl. 4: Comparison of RCAs with Hallegatte et al. (2007)

In the following Table S1 we summarize the results of the comparison between our results and climate analogues as identified by a minimum distance approach (Hallegatte et al., 2007) for some European cities. For each city the latter identifies two analogues, depending on the climate projection. We then check if they lie within the identified RCA region. The results are interpreted in the discussion section of the main article.

City	Climate Cluster	Climate Analogue Region in Hallegatte (2007)		Agree-ment A	Agree-ment B
		Weaker climate response - A	Stronger climate response – B		
Berlin	5	Italy, Reg. Rome	Spain, Reg. Salamanca	No	Yes
Copenhagen	6	France, Reg. Paris	Tirana	Yes	Yes
Helsinki	1	-	-	No	No
Paris	3	France, Reg. Bordeaux	Spain, Reg. Cordoba	No	Yes
Rome	8	-	Not on the map	No	No

Table S1: CAs from Hallegatte et al. (2007) compared with CAs gained by our cluster method.

Suppl. 5:

In Figure S3 we show the pixels where the climate change signal is small compared to the uncertainty range of the projections. In terms of our algorithm this means that the current climate of a location shares a cluster with the lower and upper bound of its climate projections.

Pixels from clusters one and two (red and ocean blue) don't show "no climate change" as defined by our cluster-oriented metrics. These are also the clusters with the smallest extend in climate space (see Table 1, variance ranks 10 and 9). Relatively large areas are found for clusters 6, 8 and 10 (cyan, yellow and magenta) in accordance with their larger variance (ranks 1, 3 and 4).

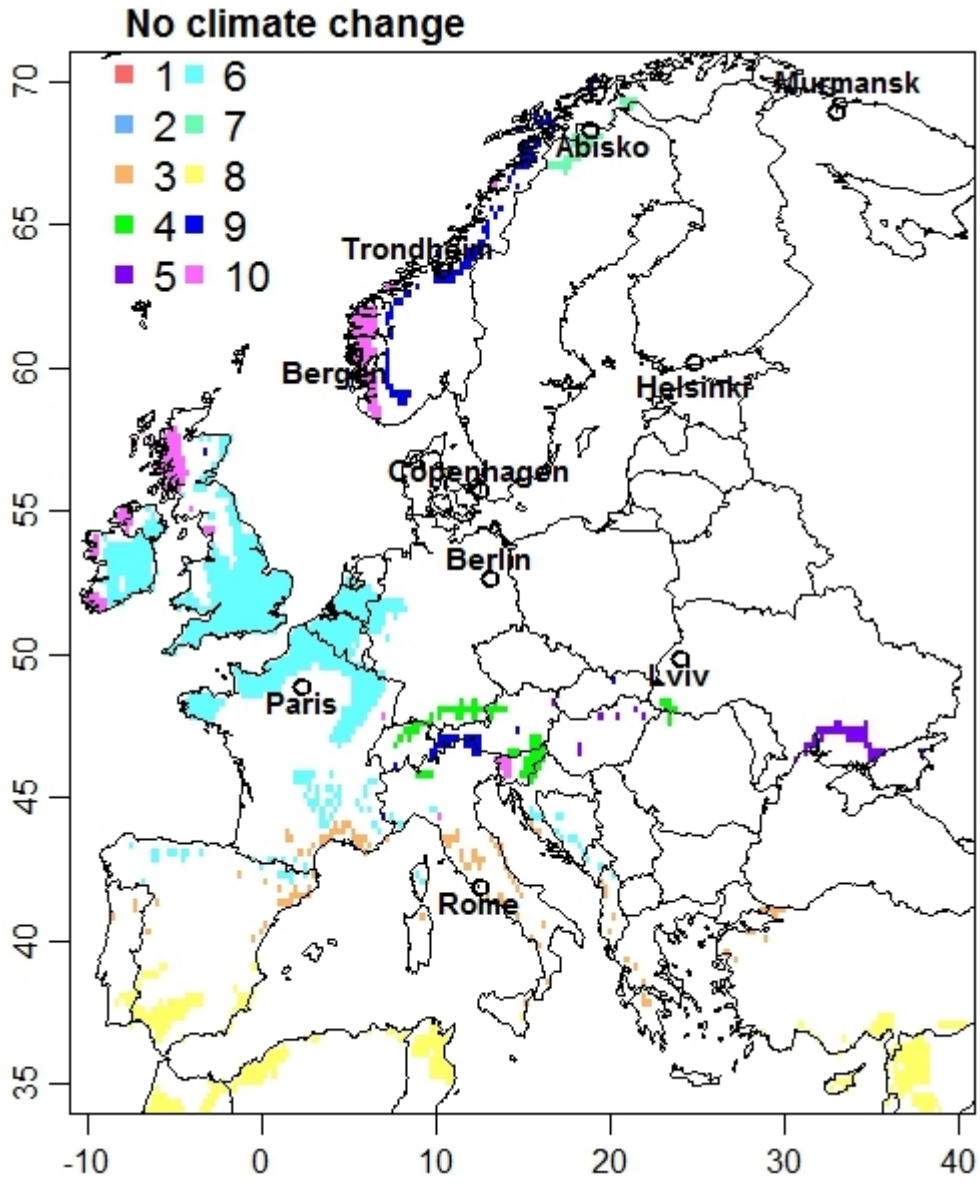

Figure S3: Pixels i where C_i, L_i and U_i share the same cluster (Colors as in Figure 2).

Literature Cited (here in the Supplement but not in the main text)

Hartigan JA, Wong MA (1979) A k-means clustering algorithm. Appl Stat 28:100–108 doi:10.2307/2346830

Hubert L, Arabie P (1985) Comparing partitions. J Classif 2:193–218 doi:10.1007/BF01908075

Kohonen T (1998) The self-organizing map. Neurocomputing 21:1–6 doi:10.1016/S0925-2312(98)00030-7

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J (eds) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, CA, p 231–297

Steinley D (2006) K-means clustering: a half-century synthesis. Br J Math Stat Psychol 59: 1–34 doi:10.1348/000711005X48266